(12) UK Patent Application (19) GB (11) 2 348 069 (13) A

(43) Date of A Publication 20.09.2000

(21) Application No 9927985.3

(22) Date of Filing 29.11.1999

(30) Priority Data
(31) 09217706 (32) 21.12.1998 (33) US

(51) INT CL[7]
H04N 7/52

(52) UK CL (Edition R)
H4F FEHM F22 F32

(56) Documents Cited
EP 0669587 A2 WO 98/24032 A1 US 5576758 A
http://www.microsoft.com/technet/winmedia/tips/asfabcs.asp
http://www.beaglesoft.com/SSManual/index.htm
http://www.academy.smc.edu/.../htmfiles/smil.htm
http://www.apple.com/quicktime/specifications.html
http://www.rockmug.org/news1.6html

(58) Field of Search
UK CL (Edition R) H4F FEHM FEHX FRG FRP FRT FRX
INT CL[7] H04N 1/41 7/30 7/32 7/34 7/36 7/50 7/52 7/54 7/56 7/62
ONLINE: WPI; JAPIO; EPODOC; COMPUTER; INTERNET

(71) Applicant(s)
International Business Machines Corporation
(Incorporated in USA - New York)
Armonk, New York 10504, United States of America

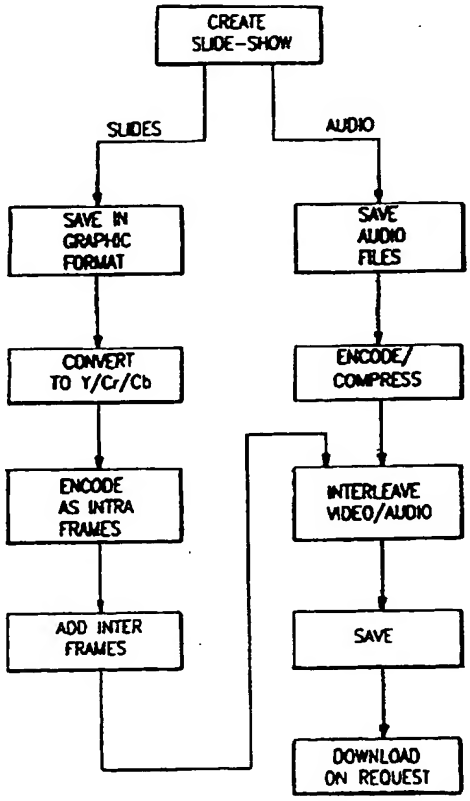
(72) Inventor(s)
Zohar Sivan

(74) Agent and/or Address for Service
C J Ling
IBM United Kingdom Limited, Hursley Park,
WINCHESTER, Hampshire, SO21 2JN,
United Kingdom

(54) Abstract Title
Representation of a slide-show as video

(57) A method for representing as a compressed video clip a slide-show including a plurality of images to be presented in sequence at respective predetermined timing intervals. Each image is encoded as a single encoded video frame. The encoded frames are arranged in a data structure corresponding to an encoded video sequence, such that at least a first one of the encoded frames is separated from a succeeding, second one of the encoded frames by a number of null frames responsive to the predetermined timing interval between the images in the sequence corresponding to the first and second frames.

FIG. 2



GB 2 348 069 A

FIG. 1

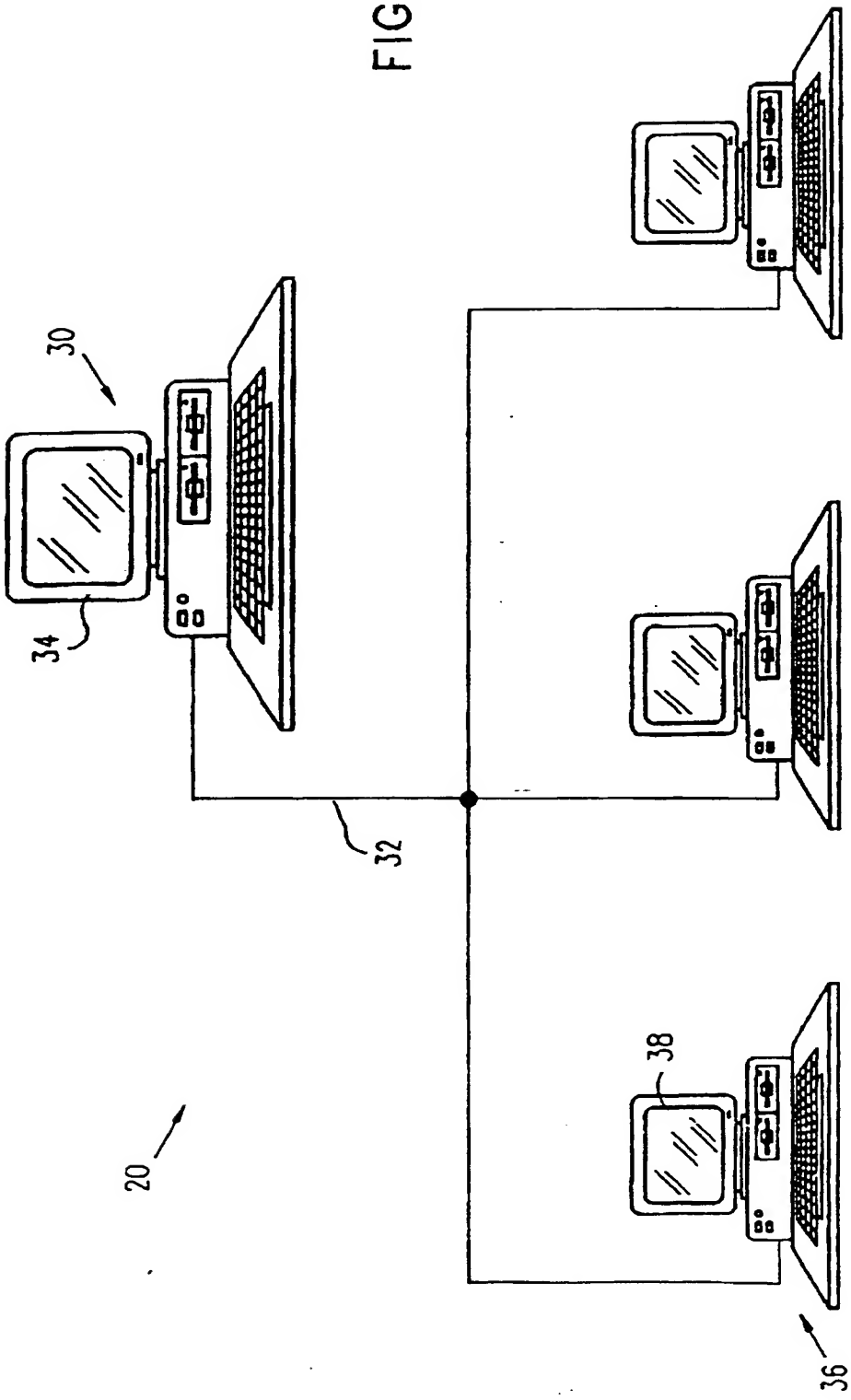


FIG. 2

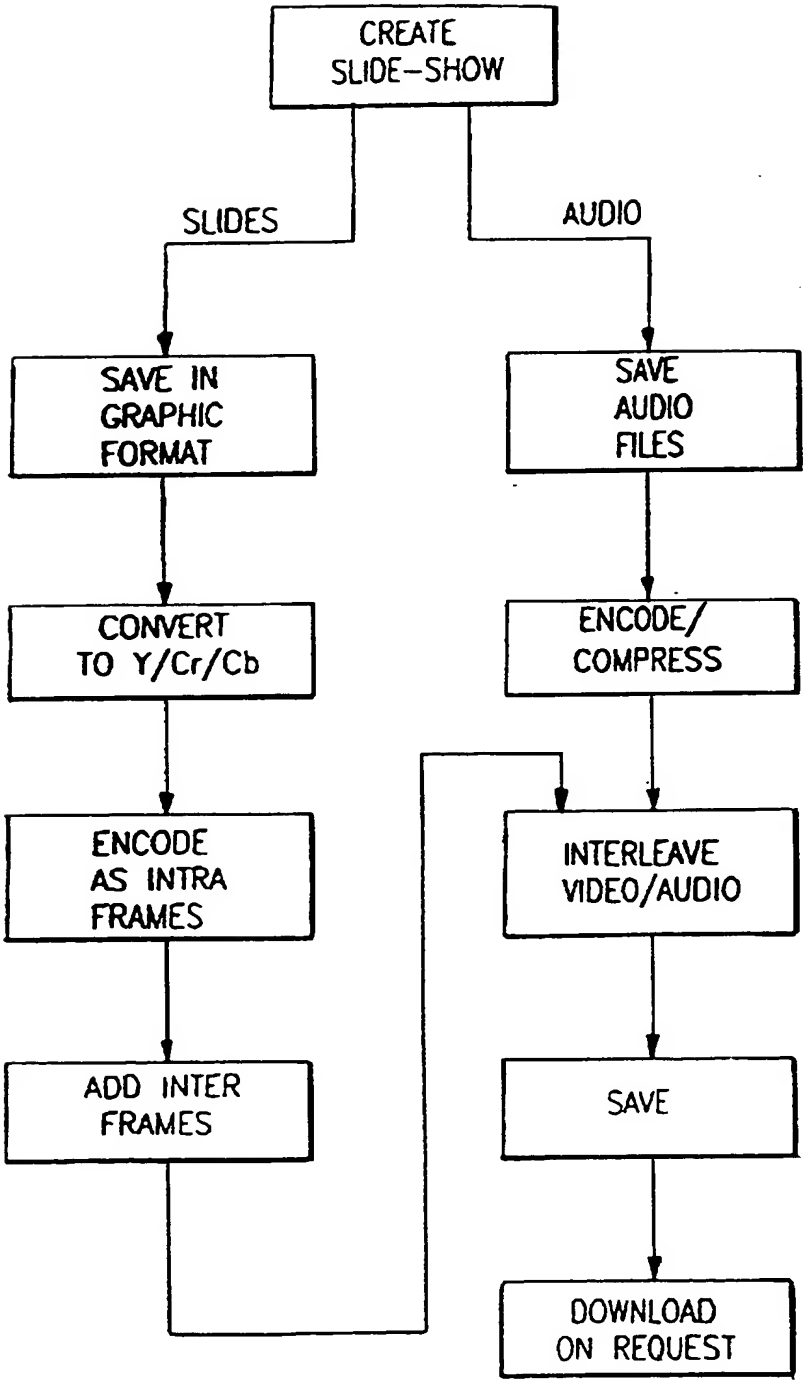


FIG. 3

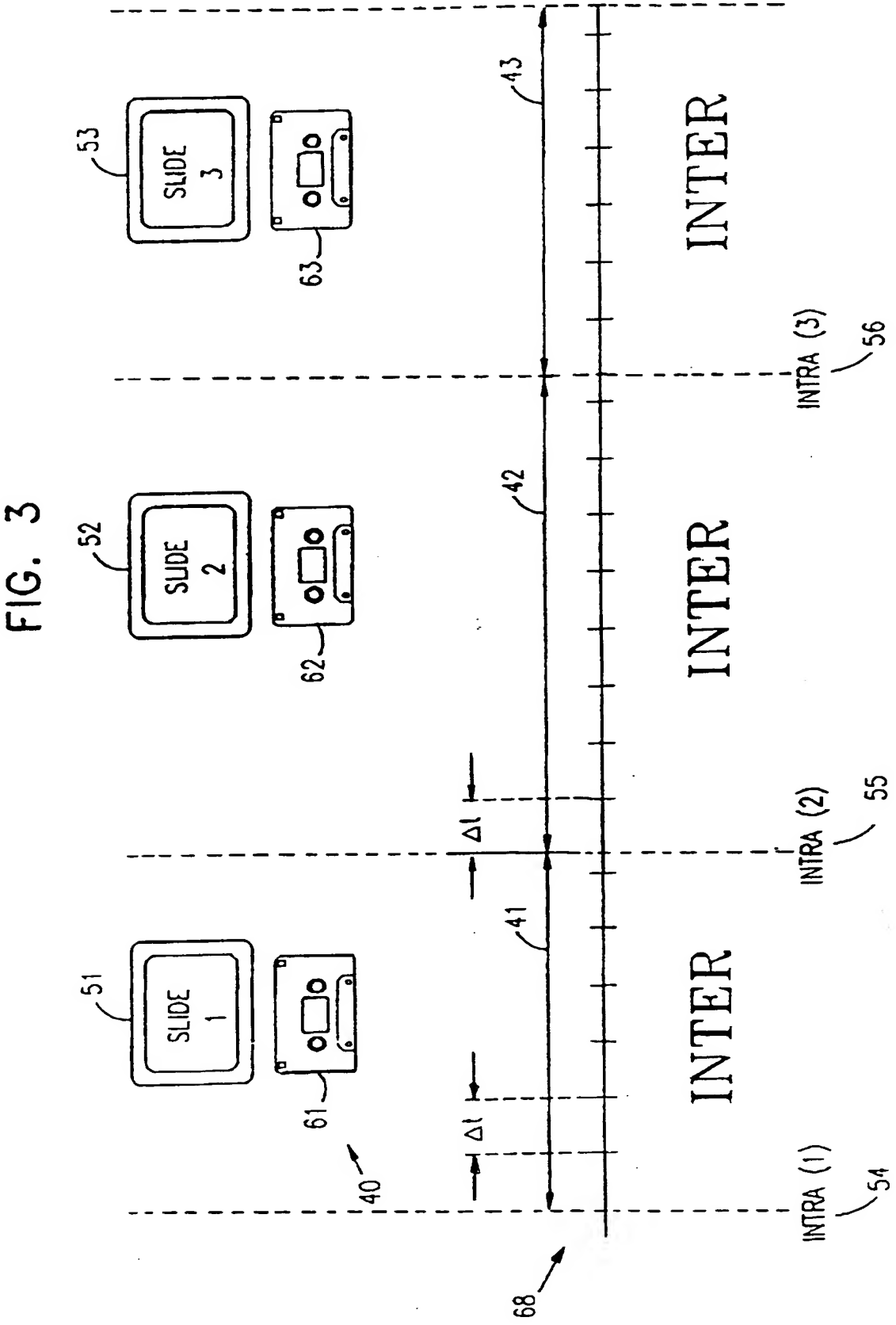


FIG. 4

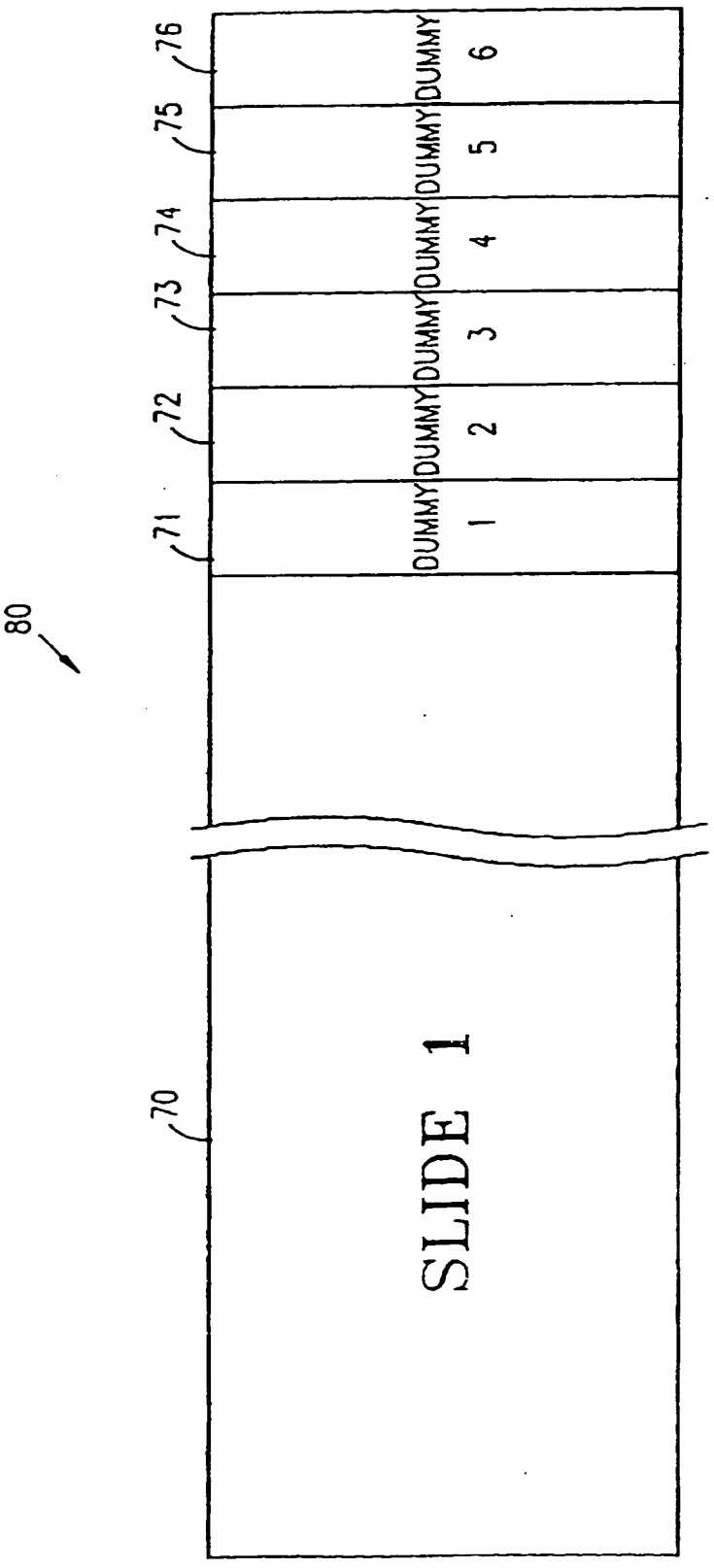
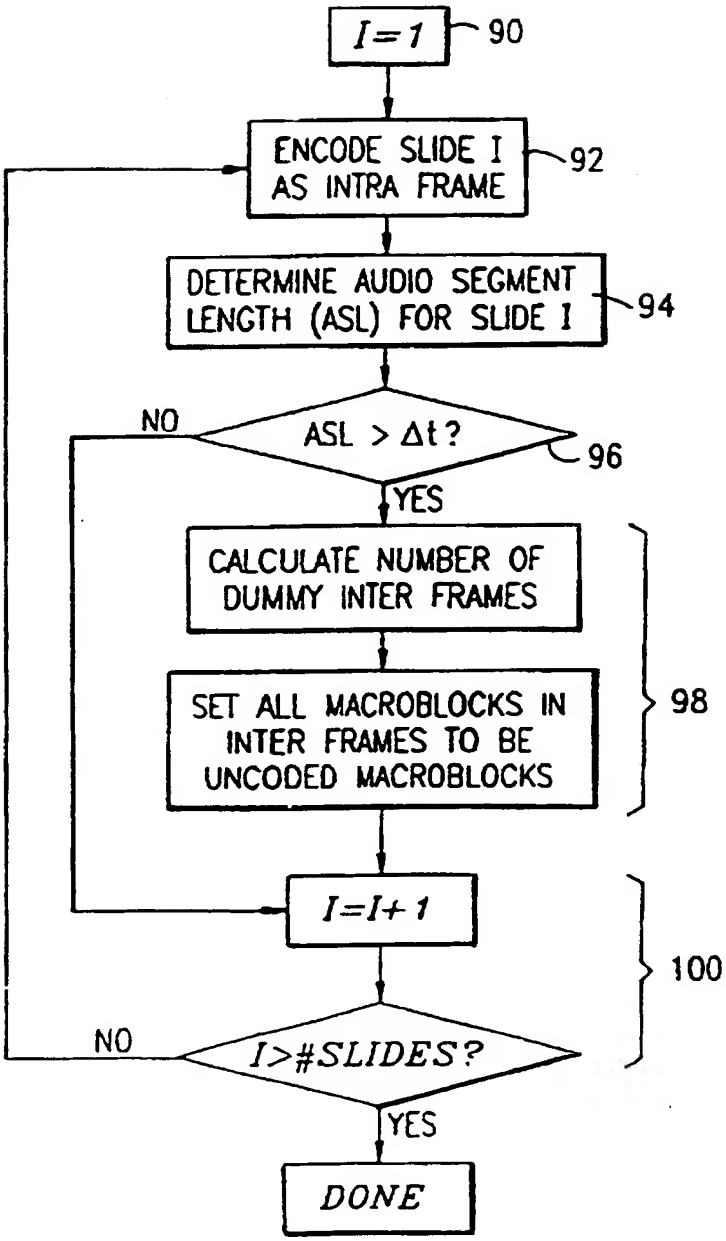


FIG. 5



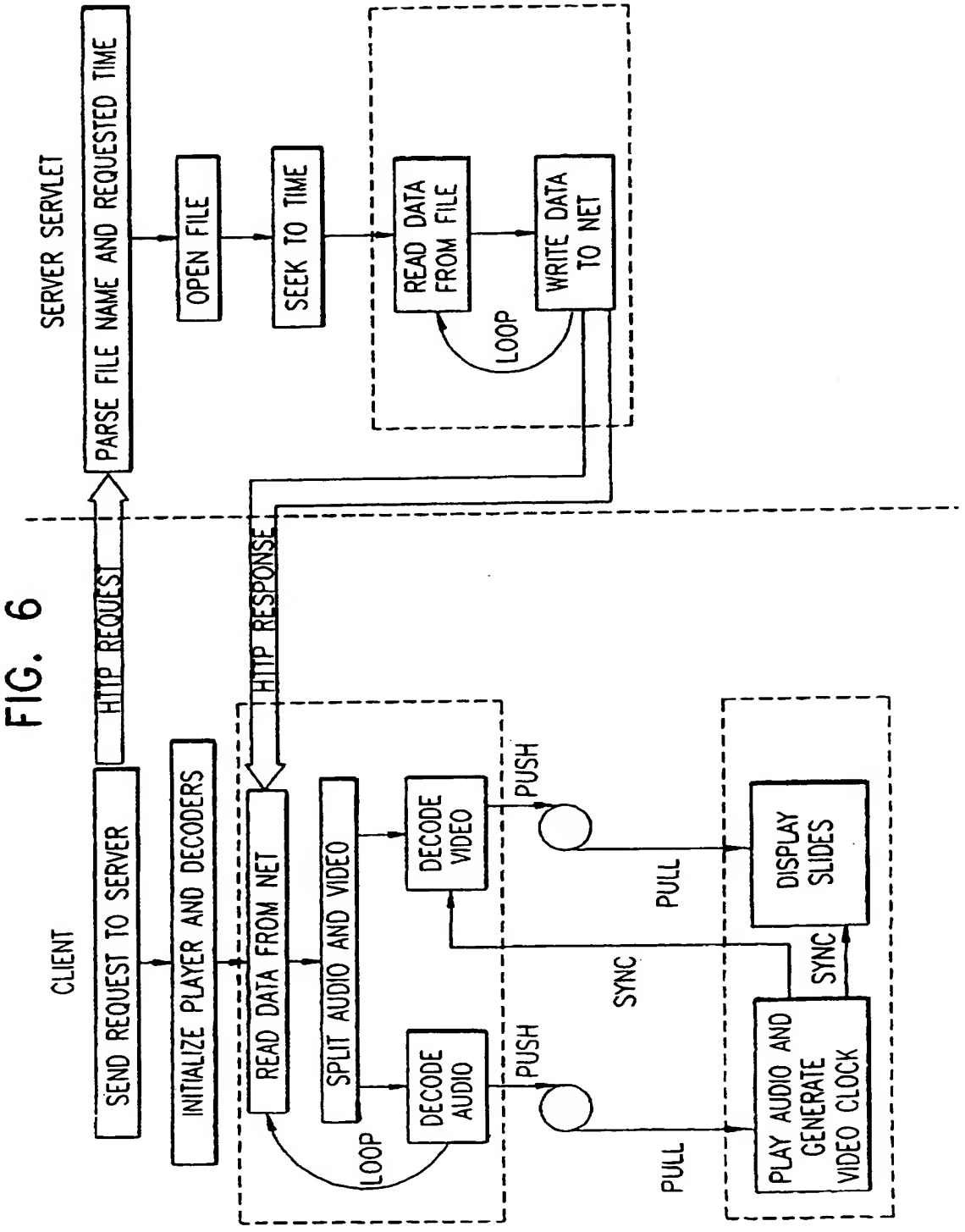


FIG. 6

REPRESENTATION OF A SLIDE-SHOW AS VIDEO

Field of the Invention

The present invention relates generally to audio-visual presentations, and specifically to methods and devices to process audio-visual media for storage, transmission and viewing.

Background of the Invention

Slide-show presentation software, including options for text, graphics, and accompanying music, is well known in the art. Application programs, such as PowerPoint97 (produced by Microsoft Inc., Redmond, Washington) and FreeLance (produced by Lotus Development Corp., Cambridge, Massachusetts), additionally allow a user to record a vocal segment of a presentation in a series of separate WAV files, each such file corresponding to a voice recording made for a respective slide (Powerpoint and Microsoft are trademarks of Microsoft Corporation and Freelance and Lotus are trademarks of Lotus Development Corporation). While it is possible to play each individual WAV file in synchronization with the respective slide during a slide-show being presented at the site of a server holding the files, viewing the presentation remotely is not possible using this software unless there is a direct connection between the server and a computer showing the slide-show.

RealNetworks (Seattle, Washington) produces RealPresenter software as a "plug-in" for PowerPoint97, which transmits a slide-show with accompanying audio by converting each slide into a JPEG image, converting the audio to RealAudio format, and interleaving these data to create a RealMedia video file. The resultant video file, however, can be downloaded only from a suitable RealNetworks server and is playable only by a user who has the associated RealNetworks video playing software.

The H.263 video encoding standard is well known in the art as a protocol for preparing a compressed video clip, which is typically transmitted across a network. Typical uses of an H.263 encoder include compressing prerecorded video for transmission and storage, as well as low bitrate video conferencing. Video sequences encoded using the H.263 coding standard typically comprise two main types of frames: an INTRA frame, representing a substantially complete encoding of an original image; and one or more INTER frames, following the INTRA frame, which represent changes in a current video frame relative to the previous frame. Compression of an original video clip with a large data volume is achieved through the use of the INTER frames, which generally have substantially smaller data volume than the corresponding INTRA frame. Details regarding the standard may be found in, for instance, the Draft ITU-T Recommendation H.263, "Video coding for very low bitrate

communication," May, 1996. Other video standards, such as the H.261 video coding standard, are also known in the art for use in generating a compressed video clip for low bandwidth transmission.

Summary of the Invention

In preferred embodiments of the present invention, a video encoder converts a computer-generated slide-show and audio segments corresponding to each slide thereof into a video clip for subsequent local playback and/or transmission to a remote site. An important concept in these embodiments is to conserve memory used by the video clip by generally having only one memory-consuming video frame represent each slide, and having essentially empty, place-holding frames during the time allotted for the audio segment corresponding to that slide. The place-holding frames are identified as place-holders to a video decoder processing the clip, so that the video frame representing the slide is viewed substantially without interruption throughout the audio segment.

In some preferred embodiments of the present invention, a video encoder is used to convert the slide-show into a video sequence substantially in accordance with the H.263 standard, in which a single H.263 INTRA frame corresponds to each slide. The INTRA frames are preferably compressed, in accordance with H.263, so as to adjust the data volume of the encoded video clip to match the bandwidth of a channel, such as an Internet connection, over which the clip is to be transmitted. Between the successive INTRA frames, the sequence includes substantially empty intermediate (INTER) frames (representing null changes from the INTRA frame in accordance with the H.263 video coding standard), which are encoded into a very small amount of memory. The number of INTER frames and their frame rate (known as a "temporal reference" in H.263) are preferably set responsive to the length of time during which the corresponding slide is to be presented, which is generally equal to the duration of the audio segment corresponding to the slide. The series of substantially empty INTER frames therefore indicates to an H.263 decoder playing the video that the INTRA frame should remain unchanged for the duration of the audio segment corresponding to the INTRA frame.

When for a given slide, the INTRA frame and an appropriate number of INTER frames have been encoded, a new INTRA frame and set of INTER frames corresponding to the next slide are encoded. This procedure continues until all of the slides in the slide-show have been processed. Thereafter, the audio segments are preferably interleaved with the video frames. In this manner, a single video file, comprising all of the slides and associated audio segments, is prepared using a minimal amount of memory.

In some preferred embodiments of the present invention, the video file thus prepared is downloaded to a remote client from a suitable server, preferably using a Hypertext Transfer Protocol (http), as is known in the art. The slide-show is viewed at the client using Java(TM)-language application programs (applets) running on the client and the server. No other application programs or "plug-ins" are required. This software configuration allows the slide-show to be streamed to the client and eliminates the need to download the entire video file before viewing. The data streaming features of the present invention are afforded by http and by the novel use of video encoding generally and of INTER and INTRA frame types specifically, as provided by common video encoding standards, such as H.263. Preferably, a user viewing the video can move forward or backward in the slide-show, or jump to a desired location therein, without opening and closing audio files corresponding to each slide.

Although preferred embodiments are described herein with reference to the H.263 standard, it will be appreciated by those skilled in the art that the principles of the present invention may similarly be applied using other methods and standards of video encoding, as well. Similarly, although preferred implementations of the present invention use http and Java application programs to download and play the computer-generated slide-show at the client, it will be understood that other network protocols and media player software programs may also be used for this purpose.

There is therefore provided, in accordance with a preferred embodiment of the present invention, a method for representing as a compressed video clip a slide-show including a plurality of images to be presented in sequence at respective predetermined timing intervals, including:

encoding each image as a single encoded video frame; and

arranging the encoded frames in a data structure corresponding to an encoded video sequence, such that at least a first one of the encoded frames is separated from a succeeding, second one of the encoded frames by a number of null frames responsive to the predetermined timing interval between the images in the sequence corresponding to the first and second frames.

Preferably, the method includes including interleaving the data structure with audio data, wherein arranging the encoded frames includes setting the number of null frames responsive to the duration of a segment of the audio data associated with the first encoded frame.

Preferably, arranging the encoded frames includes setting a frame rate of the null frames responsive to the predetermined timing interval.

In a preferred embodiment, the method includes transmitting the data structure to a remote site, where the plurality of images are presented in sequence at the respective predetermined timing intervals.

Preferably, the image in the sequence corresponding to the first frame is displayed while the second frame is being transmitted.

Preferably, transmitting the data includes transferring the data over a network using a Hypertext Transfer Protocol.

Further preferably, transmitting the data includes seeking one of the encoded frames in the sequence responsive to a request from the remote site and transmitting the data structure starting from the encoded frame that was sought.

There is further provided, in accordance with a preferred embodiment of the present invention, a method for representing as a compressed video clip a slide-show including a plurality of images to be presented in sequence, including:

encoding each image as a single compressed video frame; and

arranging the compressed frames in a data structure corresponding to a H.263 compressed video sequence.

Preferably, encoding each image includes encoding the image as an INTRA frame, and arranging the compressed frames includes separating a first INTRA frame from a succeeding INTRA frame by inserting one or more null INTER frames therebetween. Preferably, inserting the one or more null INTER frames includes setting the number of null INTER frames responsive to a predetermined timing interval between the first INTRA frame and a succeeding INTRA frame, which most preferably includes determining the duration of an audio segment corresponding to the first INTRA frame and choosing a number of null frames responsive thereto.

Preferably, inserting the one or more null INTER frames includes setting substantially all macroblocks in the INTER frames to be uncoded macroblocks.

There is also provided, in accordance with a preferred embodiment of the present invention, apparatus for transmitting a video clip representation of a slide-show including a sequence of slides, including a slide-show server, which transmits a data structure corresponding to a sequence of encoded video frames representing the sequence of the slides, and in which at least a first one of the encoded frames is separated from a succeeding, second one of the encoded frames by a number of null frames responsive to a predetermined timing interval between slides in the slide-show corresponding to the first and second encoded frames.

Preferably, the predetermined timing interval is set responsive to the duration of an audio segment associated with a slide in the slide-show corresponding to the first encoded frame.

In a preferred embodiment, the encoded video frames are encoded as INTRA frames according to the H.263 video coding standard, wherein the null frames are encoded as INTER frames according to the H.263 video coding standard and substantially all macroblocks in the INTER frames are set as uncoded macroblocks.

Brief Description of the Drawings

The invention will now be described, by way of example only, with reference to the accompanying drawings, in which:

Fig. 1 is a schematic illustration of a slide-show server and clients thereof, in accordance with a preferred embodiment of the present invention;

Fig. 2 is a flow chart schematically illustrating creation and transmission of a slide-show by the server of Fig. 1, in accordance with a preferred embodiment of the present invention;

Fig. 3 is a schematic timing diagram of a sequence of frames and audio in a video clip prepared in accordance with the method of Fig. 2, in accordance with a preferred embodiment of the present invention;

Fig. 4 is a schematic illustration of a data structure generated in accordance with the method of Fig. 2, in accordance with a preferred embodiment of the present invention;

Fig. 5 is a flow chart that schematically illustrates details of the method of Fig. 2, in accordance with a preferred embodiment of the present invention; and

Fig. 6 is a flow chart that schematically illustrates retrieval, transmission and playback of a slide-show, in accordance with a preferred embodiment of the present invention.

Detailed Description of Preferred Embodiments

Fig. 1 is a schematic illustration of apparatus 20 for representing a slide-show and accompanying audio segments as a video clip, the apparatus comprising a slide-show server 30 and one or more optional clients 36, in accordance with a preferred embodiment of the present invention. Server 30 preferably prepares the video clip for later viewing on a monitor 34 coupled to the server or alternatively receives

and stores the video clip after preparation thereof by one of clients 36 or by another computer. Additionally or alternatively, the clip is uploaded to an optional network 32 and subsequently downloaded by one or more clients 36, such that the clip may be watched on a monitor 38 as it is downloaded, or may be downloaded and stored for later viewing. Network 32 may comprise the Internet, a Local Area Network, or any other suitable computer network that is known in the art.

In a typical use of this embodiment of the present invention, a user of client 36 uses a browser program to access a Web page maintained by slide-show server 30 and selects a slide-show presentation he would like to view. Server 30 transmits the file, which is viewed in real-time by the user without being first downloaded in its entirety, i.e., the file is received and played by client 36 in the form of streaming media. Preferably, both server 30 and client 36 run standard operating systems, such as Microsoft Windows NT and Windows 95, respectively, and the slide-show is downloaded from the server to the client using the Hypertext Transfer Protocol (http), as is known in the art. To download and play the slide-show, the server and client preferably use suitable Java (TM) language application programs (a "servlet" and an applet, respectively), which are described further hereinbelow, without the need for additional dedicated server software or browser plug-ins. Unless the user so selects, the file is typically not stored in client 36 for repeat viewing.

Fig. 2 is a flow chart schematically illustrating a method for creating and transmitting a computer-generated slide-show file over network 32, in accordance with a preferred embodiment of the present invention. The slide-show is created using any suitable software package known in the art, such as Microsoft PowerPoint (TM) 97. After the slides are created, they are saved in a graphic file format, such as GIF, JPEG, or any other suitable format known in the art (Windows is a trademark of Microsoft Corporation). Preferably, some or all of the slides are accompanied by corresponding audio segments, which are recorded and stored as audio data. The audio data are encoded and compressed using any appropriate audio coding standard known in the art, such as G.723 (promulgated by ITU), which is most suitable for voice data, and are stored along with the slides.

Typically, the slides are generated and stored by the slide-show software package in an RGB format. In order to encode the slides as video frames, the graphic files are first converted to Y/Cr/Cb format, as is known in the video art. Then each slide is encoded, preferably as a H.263 INTRA frame, wherein quantization steps used in the encoding are adjusted, as is known in the art, to control compression of the INTRA frame responsive to the available data storage volume and/or transmission bandwidth. Preferably, the quantization steps (or other

compression-controlling parameters) are adjusted to give the best possible image quality within the limits of the given data volume and/or bandwidth. In this respect, video encoding, such as H.263 or MPEG, is superior to still picture encoding standards, such as GIF, commonly supported and used in computer-generated slide-shows, which are not adapted for adjustment to meet bandwidth constraints.

After creating the INTRA frames, a number of H.263 INTER frames are typically inserted between each pair of successive INTRA frames to create a complete H.263 video clip. Insertion of the INTER frames is described further hereinbelow with reference to the figures that follow.

When all of the slides have been encoded as INTRA frames and the appropriate number of dummy INTER frames have been inserted after the INTRA frames, as described hereinabove, the audio segments of the presentation are interleaved with the video portion. This can be performed using standard tools and file formats, as are known in the art. For example, QuickTime™ (produced by Apple Computer Co., Cupertino, California) and Microsoft's ASF have suitable file formats (Quicktime is a trademark of Apple Computer Corporation). The interleaved data can then be saved and downloaded upon request over the network from server 30 to client 36. The data are preferably downloaded and played by the client using http and suitable Java™ application programs, as mentioned hereinabove.

Fig. 3 is a schematic illustration of a timing sequence 68 of a video clip prepared by or for slide server 30, based on a sample slide-show 40, comprising first, second, and third slides 51, 52, and 53, and corresponding audio segments 61, 62, and 63, in accordance with a preferred embodiment of the present invention. Slides are preferably prepared using Microsoft's PowerPoint™ 97 or another, similar application program, and typically comprise text and/or graphics. In a typical mode of use, audio segments 61, 62, and 63 are stored in server 30 and comprise digitized audio files of speech and/or music corresponding to text or graphics on slides 51, 52, and 53. The audio can be encoded using any appropriate audio encoder suitable for the content, as is known in the art (e.g., an encoder for speech only, or for speech plus music).

Audio segments 61, 62, and 63 are characterized by respective segment lengths 41, 42, and 43 indicated by corresponding arrows in Fig. 3. Segment length 41 corresponds to, for example, the length of time that a lecturer spoke while referring to slide 51 before switching to slide 52. Additionally or alternatively, some or all of the audio segments and segment lengths do not contain information about an audio component per se, but serve only as indicators of the length of time a given slide should be played. For example, a silent slide-show might

present each slide for 30 seconds, and this would be indicated by the values of segment lengths 41, 42, and 43.

In this preferred embodiment of the present invention, the video clip is prepared using the H.263 video encoder, such that upon playback, the video will show slides 51, 52, and 53, each for lengths of time roughly corresponding to segment lengths 41, 42, and 43, respectively. A person skilled in the art will appreciate that the disclosed method of representing slides as video can also be carried out, mutatis mutandis, using the H.261 standard or other suitable video coding standards.

Video sequences encoded using the H.263 coding standard typically comprise two types of frames: an INTRA frame, representing a substantially complete encoding of an original image; and, if necessary, one or more INTER frames, following the INTRA frame, which represent changes in a current video frame relative to the previous frame. In the embodiment of the present invention shown in Fig. 3, slide-show 40 is represented by a series of INTRA frames 54, 55 and 56, each of which substantially completely encodes a corresponding slide in the slide-show. Following each INTRA frame, sequence 68 includes a series of substantially empty, or null, INTER frames, which indicate that no change is to be made in the video image displayed while the audio segment corresponding to the INTRA frame is played.

The H.263 standard allows variable time gaps between consecutive frames. The time gaps are varied by appropriately setting the "temporal reference" parameter defined by the standard. Increasing the time gaps helps to reduce the bandwidth required for transmission of the video clip, since it reduces the number of frames to be transmitted in a given time period. This feature, which is not available in other video encoding standards, such as MPEG, is particularly useful in the context of the present invention, in which the information-carrying INTRA frames are in any case widely spaced in time.

A constraint imposed by the H.263 standard is the requirement of setting the time (Δt) between any frame and the following frame to be no greater than approximately 8.54 seconds. This constraint stems from the fact that the temporal reference parameter defined by H.263 is an 8-bit number, with a range of 1-255, and the time gap is set in units of 1/29.97 sec. Setting Δt to be about 8.5 seconds between most of the frames is appropriate for most applications of the present invention. Smaller values of Δt can, of course, also be used. On the other hand, a larger maximum value of Δt may also be used if supported by a suitable video encoding standard or algorithm for use in place of H.263.

In particular, the value of Δt between an INTER frame and a following INTRA frame is often smaller than 8.5 seconds. For typical

slide-shows, in which the average duration of presentation of a slide is measured in tens of seconds or in minutes, therefore, a plurality of INTER frames are usually encoded following each INTRA frame. Accordingly, slides 51, 52, and 53 are encoded using the H.263 standard into INTRA frames INTRA(1) 54, INTRA(2) 55, and INTRA(3) 56, respectively. Following each INTRA frame are a plurality of substantially empty INTER frames 66, which are prepared as described hereinbelow. It is understood that for shorter audio segment lengths than those shown, there could be only one INTER frame, or similarly none at all.

Fig. 4 is a schematic illustration of a data structure 80 generated by or for server 30, to represent slide 51 and to include the substantially empty INTER frames corresponding thereto, in accordance with a preferred embodiment of the present invention. Data structure 80 comprises a data array 70 representing INTRA(1) 54, which was encoded from slide 51. For typical CIF (Common Intermediate Format) resolution images (352x288 pixels), approximately 100kb (kilobits) may be allotted for array 70, although in some applications, a greater or smaller memory allocation may be appropriate. Responsive to audio segment length 41, a total of six dummy arrays 71, 72, 73, 74, 75, and 76, representing the substantially empty INTER frames, are encoded, each INTER frame typically requiring approximately 440 bits of data (for CIF resolution images). The H.263 standard also supports several other image sizes which may be used in other embodiments of the present invention, including sub-qcif (128x96 pixels), qcif (176x144 pixels), 4cif (704x576 pixels), and 16cif (1408x1152 pixels). Other image sizes may also be used, although preferably the sizes are of dimensions (in pixels) that are evenly divisible by 16, for convenience in block encoding, as is used in H.263, MPEG and other video encoding methods known in the art.

Using the method described herein of H.263 encoding of a slide-show, a video clip is created that typically does not require a significantly greater amount of memory than that required for the slides alone. For example, a 100kb CIF resolution slide that is to be shown for one minute requires seven dummy INTER frames to follow it (setting Δt between most of the frames to be 8.5 seconds). If each INTER frame uses 440 bits of memory, then the total amount of memory utilized to represent the slide as a one minute video segment is approximately 103kb, not including the audio portion. Additionally, in this example, the 103kb representing the one minute video portion of the slide could be transmitted by server 30, assuming a moderate transmission speed of 16 kbps, in under seven seconds. This is within the capabilities of a typical 28.8 kbps modem, and leaves a substantial time period for transmitting the corresponding audio portion and any necessary audio/video interleaving data.

Fig. 5 is a flow chart that schematically illustrates the operation of server 30 in preparing a video clip, in accordance with a preferred embodiment of the present invention. Fig. 5 describes in detail the steps of encoding slides as INTRA frames and adding INTER frames included in the method of Fig. 2. In step 90, a slide counter variable I is initialized to 1. Slide I is encoded as an INTRA frame in step 92, and the length of the corresponding audio segment is determined in step 94. In step 96, a test is performed to determine whether the audio segment length is greater than Δt, in which case program control will be passed to steps 98 in order to create and insert dummy INTER frames. If, alternatively, the audio segment length is not greater than Δt, then I is incremented in steps 100, and, if there remain more slides to encode, control is passed back to step 92 to encode the next slide.

In steps 98, the number of dummy INTER frames to insert is preferably calculated by dividing the audio segment length corresponding to slide I by the value of Δt between most frames (e.g., 8.5 seconds) and rounding the result down, if necessary, to the nearest integer. If rounding down is performed, then the value of Δt between the last INTER frame and the INTRA frame representing slide I + 1 is preferably set to a smaller value, Dtreduced. This smaller value is preferably set such that the sum of all of the Δt's for slide I together with Dtreduced is substantially equal to the audio segment length corresponding to slide I. Alternatively, the number of INTER frames is set by dividing the audio segment length corresponding to slide I by Δt and rounding up to the nearest integer. By rounding up, an additional INTER frame is often used compared to the number that would be used by rounding down, but the value of Δt is maintained substantially constant. Further alternatively, Δt for all of the INTER frames between two successive INTRA frames is chosen so as to divide evenly the audio segment length corresponding to the first of the two INTRA frames, so that substantially no rounding is required at the last INTER frame. In either case, the INTER frames are coded as dummy frames, preferably by setting all macroblocks in the frame to be uncoded macroblocks (as defined by the H.263 coding standard). Subsequently, I is incremented, and if there remain further slides to encode, control is passed back to step 92.

Although it is assumed in the preferred embodiment described hereinabove that Δt is chosen so as to minimize the number of INTER frames that must be interposed between the successive INTRA frames, such a choice of Δt is not a necessary constraint. In some cases, it may be desirable to use a smaller value of Δt and thereby increase the number of INTER frames between successive INTRA frames, for example, so as to reduce the impact of lost packets on an Internet transmission.

When all of the slides have been encoded as INTRA frames and the appropriate number of dummy INTER frames have been inserted after the

INTRA frames, as described hereinabove, the audio segments of the presentation are interleaved with the video portion. Although such interleaving may be performed using standard tools and file formats, such as QuickTime (TM) or ASF, as mentioned hereinabove, Table I below illustrates an exemplary file format known as HAV (Haifa Research Laboratory Audio Video), developed by IBM Corporation and covered by copyright belonging to IBM, which is particularly useful for this purpose:

TABLE I

```
// Format of the fixed-size Macro Frame Header which
// is present in every Macro Frame.
// All WORD parameters in *network* byte order (big-
// endian).
// All DWORD parameters in *network* byte order.
// The initial macro block will not be byte stuffed.
#pragma pack(push, 1)
struct MacroFrame {
    DWORD dwSyncCode;          // All headers are
                               // preceded by this code
    BYTE byFlags;              // Flags.
        // Bit 0 = 1 always.  Bit 1 = 1 if this
        // frame has stuffing.  Bit 2 = 0 if this
        // frame includes an INTRA video frame.
        // Otherwise it is set to  1. Bits 3-7
        // indicate the number of stuffed bytes
        // in this frame.
    BYTE byVersion;            // Version number
    BYTE byHiSizeVideo;        // Upper 8 bits
    WORD wLoSizeVideo;         // How many bytes of video
                               // in this frame
    WORD wSizeAudio;           // How many bytes of audio
                               // in this frame
    WORD wSizeUser;            // How many bytes follow
                               // this structure
    BYTE byHiSequence;         // Upper 8 bits
    WORD wLoSequence;          // Position in stream of
                               // this frame, in ms
    WORD wLenSequence;         // How much time should
                               // this block play for,
                               // in ms
};
    // Macro Frame header will be 18 bytes long.
    // Using 24 bits for video size allows up to 4
    // video frames of 1000x1000 RGB.
    // Using 24 bits for sequence allows clips up to
            // 4.6 hours (279 minutes) long.
    // Using 16 bits for audio size allows up to 0.3
```

```
    // seconds of uncompressed CD audio.
    // Using 16 bits for LenSequence allows up to 65
    // seconds in a macro frame.

// Format of the PlayParams block
struct PlayParams {
  BYTE  Tag;                    // Always 0
  DWORD dwVideoSpeed;           // How many time units for
          // each video ref #.  In 100ns units.
          // (reciprocal of frame rate).
          // 0xffffffff for audio only.
  FOURCC fccVideo;              // Video format. 0 = audio
                                // only
  FOURCC fccAudio;              // Audio format. 0 = video
                                // only
  WORD  wFrameRate;             // FPS target
  BYTE  byRateVideo;            // upper 8 bits
  WORD  wLoRateVideo;           // BPS max, e.g. 20kbps =
                                // 20480.  0 = audio only.
  BYTE  byRateAudio;            // upper 8 bits
  WORD  wLoRateAudio;           // BPS max, e.g. 6.4kbps =
                                // 6554.  0 = video only.
  WORD  wRateExtra;             // BPS max for non-
        // audio/video data
        // (frame headers and params)
  BYTE  byHiLength;             // Upper 8 bits
  WORD  wLoLength;              // Length of clip, ms
  WORD  wWidth;                 // Video Pixels, X.  0 for
                                // audio only.
  WORD  wHeight;                // Video Pixels, Y.  0 for
                                // audio only.
  BYTE  byParamFlags;           // Flags.  Currently, bit
                                // 0 = 1 if this frame
                                // has no rate control
  char  szTitle[128];           // Null-terminated
                                // C strings
  char  szAuthor[128];
  char  szDescription[128];
  char  szCopyright[128];
};
#pragma pack(pop)
```

As illustrated by the table above, a HAV file includes two or more "Macro Frames" which appear concatenated together, as a stream of bytes (octets). A Macro Frame has up to four parts, which are always in the same order: an 18 byte Macro Frame Header, optional Play Parameters, optional video, and optional audio. The space allocated in the Macro

optional video, and optional audio. The space allocated in the Macro Frame for the Play Parameters, Video Size and Audio Size is given in the Macro Frame Header. Typically, when HAV files are stored on a disk, the first Macro Frame has the Play Parameters, while all subsequent Macro Frames do not have the Play Parameters. Alternatively, both the first Macro Frame and Macro Frames subsequent thereto have the Play Parameters.

By convention, the PlayParams block (the representation of the Play Parameters) always begins with a "00" byte. If, during playback, a PlayParam beginning with a non-"00" byte is encountered, the playback should ignore the PlayParam.

The first Macro Frame, unlike all subsequent Macro Frames, is never "byte stuffed." Byte stuffing means that in each Macro Frame, all "00 00 f0" hex patterns other than the initial pattern at the beginning of the Macro Frame are replaced with "00 00 f0 ff." During playback, if "00 00 f0 ff" is encountered in the stream it is "unstuffed" and replaced with "00 00 f0." This allows a playback parsing routine to locate the beginning of Macro Frames if the stream is played from arbitrary positions, and allows the playback parsing routine to recover from missing or out-of-sequence data when the stream is transmitted over a non-error correcting protocol connection.

FOURCC is a RIFF (Resource Interchange File Format) convention, known in the art, for identifying video and audio file formats. The current HAV implementation supports the H.263 video standard, the G.723 audio standard, and an IBM® proprietary audio compression algorithm which is currently known as MRTC and is based on the G.723 standard.

Fig. 6 is a flow chart that schematically illustrates retrieval, transmission and playback of a slide-show, in accordance with a preferred embodiment of the present invention. To enable playback of the video clip, client 36 should be equipped with a video player that supports the chosen interleaving file format, such as HAV, and can decode the chosen audio coder, as well as the H.263 video coding standard. The request for and downloading of the video clip, preferably using http, as described hereinabove, and the playback operation of client 36 are preferably controlled by a Java® applet, which may be downloaded from server 30. Similarly, uploading of the video clip from the server to network 32 is preferably controlled by a Java® program running on the server, known as a "servlet." The use of Java® programs to download and display the slide-show has the advantages of being platform-independent and requiring no browser plug-ins or additional software packages for support.

Using the Java® applet, or a suitable substitute, client 36, or any other appropriately configured computer, can display the video clip, perform seeking operations to any desired slide or time-point in the

presentation, or execute any other functions with respect to the video clip that are commonly associated with computer-generated slide-show viewing. The seeking operations are facilitated by the use of a separate INTRA frame to represent each slide. When a user of client 36 inputs a request to jump to a particular point in the slide-show, the applet running on the client sends a corresponding seek request to server 30. The servlet running on the server finds the nearest INTRA frame which precedes the point indicated by the user and transmits that frame (or a sequence of frames beginning with that INTRA frame) to the client. The INTRA frame can be decoded by the client substantially without the need for any additional frames or information. This independence of display of each of the slides in the slide-show would not be possible if some of the slides were represented as INTER frames. On the other hand, it is possible and may in some circumstances be desirable to represent some of the slides as INTER frames, so as to reduce the overall data volume of the slide-show.

A user of client 36, running the applet downloaded from server 30, typically selects a computer-generated slide-show he would like to view, and possibly a desired time-point or slide number therein. Responsive to the selection, client 36 sends a request to server 30, preferably using http. The servlet running on server 30 preferably parses the file name and optional time-point, accesses the file, and begins to send data representing the desired slide-show on network 32, beginning at the time-point if one is selected.

After initializing the player and decoders, client 36 downloads the data from the network, and splits and decodes the audio and video segments contained in the downloaded data. The audio and video segments are preferably stored in separate FIFO (first in, first out) cyclic buffers, where they are maintained until transferred therefrom to be played or displayed by an audio player or a display unit, respectively. As described hereinabove, a video frame is displayed for the duration of the audio segment corresponding thereto. Responsive to a timing control signal generated responsive to the audio segment, the video frame is replaced by a subsequent video frame. In this manner, the applet running on client 36 synchronizes the display of each slide with the appropriate audio segment. As each successive video frame is displayed, it frees space in the cyclic video buffer, which had been used to store the frame, to be allocated for other, not-yet-decoded, video frames. Decoding of a video frame is preferably initiated by detection of sufficient free memory space in the video cyclic buffer.

Although preferred embodiments of the present invention are described hereinabove with reference to the H.263 standard, it will be understood that the principles of the present invention may similarly be applied using any other suitable method of video encoding, such as MPEG,

H.261 or other methods known in the art. H.263 is believed to offer the most advantageous implementation of the present invention among common video encoding methods that are currently known in the art. Some of the advantages of H.263 are described hereinabove. It will be appreciated, however, that the preferred embodiments described above are cited by way of example, and the full scope of the invention is limited only by the claims.

CLAIMS

1. A method for representing as a compressed video clip a slide-show comprising a plurality of images to be presented in sequence at respective predetermined timing intervals, the method comprising:

encoding each image as a single encoded video frame; and

arranging the encoded frames in a data structure corresponding to an encoded video sequence, such that at least a first one of the encoded frames is separated from a succeeding, second one of the encoded frames by a number of null frames responsive to the predetermined timing interval between the images in the sequence corresponding to the first and second frames.

2. A method as claimed in claim 1, further comprising interleaving the data structure with audio data.

3. A method as claimed in claim 2, wherein the step of arranging the encoded frames comprises setting the number of null frames responsive to the duration of a segment of the audio data associated with the first encoded frame.

4. A method as claimed in claim 1, wherein the step of encoding each image and arranging the encoded frames comprise arranging and encoding substantially in accordance with a H.263 video standard.

5. A method as claimed in claim 1, wherein the step of arranging the encoded frames comprises setting a frame rate of the null frames responsive to the predetermined timing interval.

6. A method as claimed in claim 1, further comprising transmitting the data structure to a remote site, where the plurality of images are presented in sequence at the respective predetermined timing intervals.

7. A method as claimed in claim 6, further comprising displaying the image in the sequence corresponding to the first frame while the second frame is being transmitted.

8. A method as claimed in claim 6, wherein the step of transmitting the data comprises transferring the data over a network using a Hypertext Transfer Protocol.

9. A method as claimed in claim 6, wherein the step of transmitting the data comprises seeking one of the encoded frames in the sequence responsive to a request from the remote site and transmitting the data structure starting from the encoded frame that was sought.

10. A method for representing as a compressed video clip a slide-show comprising a plurality of images to be presented in sequence, the method comprising:

encoding each image as a single compressed video frame; and

arranging the compressed frames in a data structure corresponding to a H.263 compressed video sequence.

11. A method as claimed in claim 10, wherein the step of encoding each image comprises encoding the image as an INTRA frame.

12. A method as claimed in claim 11, wherein the step of arranging the compressed frames comprises separating a first INTRA frame from a succeeding INTRA frame by inserting one or more null INTER frames therebetween.

13. A method as claimed in claim 12, wherein the step of inserting the one or more null INTER frames comprises setting the number of null INTER frames responsive to a predetermined timing interval between the first INTRA frame and a succeeding INTRA frame.

14. A method as claimed in claim 13, wherein the step of setting the number of null INTER frames comprises determining the duration of an audio segment corresponding to the first INTRA frame and choosing a number of null frames responsive thereto.

15. A method as claimed in claim 12, wherein the step of inserting the one or more null INTER frames comprises setting substantially all macroblocks in the INTER frames to be uncoded macroblocks.

16. A method as claimed in claim 10, further comprising interleaving an audio portion with the compressed video sequence.

17. Apparatus for transmitting a video clip representation of a slide-show including a sequence of slides, comprising a slide-show server, which transmits a data structure corresponding to a sequence of encoded video frames representing the sequence of the slides, and in which at least a first one of the encoded frames is separated from a succeeding, second one of the encoded frames by a number of null frames responsive to a predetermined timing interval between slides in the slide-show corresponding to the first and second encoded frames.

18. Apparatus as claimed in claim 17, wherein the predetermined timing interval is set responsive to the duration of an audio segment associated with a slide in the slide-show corresponding to the first encoded frame.

19. Apparatus as claimed in claim 17, wherein the encoded video frames are encoded as INTRA frames according to the H.263 video coding standard.

20. Apparatus as claimed in claim 19, wherein the null frames are encoded as INTER frames according to the H.263 video coding standard and substantially all macroblocks in the INTER frames are set as uncoded macroblocks.

The Patent Office
INVESTOR IN PEOPLE

Application No:	GB 9927985.3	**Examiner:**	Frank D. Moeschler
Claims searched:	10-16	**Date of search:**	14 July 2000

Patents Act 1977
Further Search Report under Section 17

Databases searched:

UK Patent Office collections, including GB, EP, WO & US patent specifications, in:

UK Cl (Ed.R): H4F (FEHM, FEHX, FRG, FRT, FRP, FRX)

Int Cl (Ed.7): G06F-3/033; H04N-1/41, 7/30, 7/32, 7/34, 7/36, 7/50, 7/52, 7/54, 7/56, 7/62

Other: Online: WPI; JAPIO: EPODOC; COMPUTER; INSPEC

Documents considered to be relevant:

Category	Identity of document and relevant passage	Relevant to claims
X	EP 0669587 A2 (AT&T) See whole document	10-16
X	US 5576758 (ARAI et al) See Cols 4-7	10-15
X,P	http://www.microsoft.com/technet/winmedia/tips/asfabcs.asp	10-16
X	http://www.apple.com/quicktime/specifications.html	10-16
A	http://www.rockmug.org/news1.6html	

X	Document indicating lack of novelty or inventive step	A	Document indicating technological background and/or state of the art.
Y	Document indicating lack of inventive step if combined with one or more other documents of same category.	P	Document published on or after the declared priority date but before the filing date of this invention.
&	Member of the same patent family	E	Patent document published on or after, but with priority date earlier than, the filing date of this application.

BNSDOCID: <GB_____2348069A__I_>

Application No: GB 9927985.3 **Examiner:** Frank D. Moeschler
Claims searched: 1-9; 17-20 **Date of search:** 12 June 2000

Patents Act 1977
Search Report under Section 17

Databases searched:

UK Patent Office collections, including GB, EP, WO & US patent specifications, in:

UK Cl (Ed.R): H4F (FEHM, FEHX, FRG, FRT, FRP, FRX)

Int Cl (Ed.7): G06F-3/033; H04N-1/41, 7/30, 7/32, 7/34, 7/36, 7/50, 7/52, 7/54, 7/56, 7/62

Other: WPI; JAPIO; EPODOC; INSPEC; COMPUTER; INTERNET

Documents considered to be relevant:

Category	Identity of document and relevant passage	Relevant to claims
X	EP 0669587 A2 (AT&T) See whole document	1-9,17-19
X	WO 98/24032 A1 (AMERICA ONLINE) See Fig 6	1-3,5,17, 18
X	US 5576758 (ARAI et al) See Col 4 Lines 38-44	1,4,5
P, X	http://www.microsoft.com/technet/winmedia/tips/asfabcs.asp	1-3, 5-8, 17,18
P, X	http://www.beaglesoft.com/SSManual/index.htm	1-3, 5-8, 17,18
X	http://www.academy.smc.edu/vanallen/ProdGuideHTML/htm files/smil.htm	1-3, 5-8, 17,18

X	Document indicating lack of novelty or inventive step	A	Document indicating technological background and/or state of the art.
Y	Document indicating lack of inventive step if combined with one or more other documents of same category.	P	Document published on or after the declared priority date but before the filing date of this invention.
&	Member of the same patent family	E	Patent document published on or after, but with priority date earlier than, the filing date of this application.

THIS PAGE BLANK (USPTO)